Numer Algor (2008) 49:85–104
DOI 10.1007/s11075-008-9163-1

ORIGINAL PAPER

Error estimates for linear systems with applications to regularization

C. Brezinski · G. Rodriguez · S. Seatzu

Received: 2 July 2007 / Accepted: 4 January 2008 /
Published online: 1 February 2008

Abstract In this paper, we discuss several (old and new) estimates for the norm of the error in the solution of systems of linear equations, and we study their properties. Then, these estimates are used for approximating the optimal value of the regularization parameter in Tikhonov's method for ill-conditioned systems. They are also used as a stopping criterion in iterative methods, such as the conjugate gradient algorithm, which have a regularizing effect. Several numerical experiments and comparisons with other procedures show the effectiveness of our estimates.

Keywords Ill-conditioned linear systems · Regularization · Error estimates

1 Introduction

The computation of the solution of an ill-conditioned system of linear equations is a difficult numerical problem which requires special techniques. In particular, when using Tikhonov's regularization, a value of the parameter has

This work was supported by MIUR under the PRIN grant no. 2006017542-003, and the University of Cagliari.

C. Brezinski
Laboratoire Paul Painlevé, UMR CNRS 8524, Université des Sciences
et Technologies de Lille, 59655 Villeneuve d'Ascq Cedex, France
e-mail: Claude.Brezinski@univ-lille1.fr

G. Rodriguez (✉) · S. Seatzu
Dipartimento di Matematica e Informatica, Università di Cagliari,
Viale Merello 92, 09123 Cagliari, Italy
e-mail: rodriguez@unica.it

S. Seatzu
e-mail: seatzu@unica.it

to be chosen, or, for an iterative method, a good stopping criterion has to be implemented. In both cases, the norm of the error passes through a global minimum, not necessarily unique, and this corresponds to the best value of the regularization parameter. Obviously, the notion of "best value" is relative to the particular norm used to measure the error. We will adopt, as it is usual, the Euclidean norm.

Two of the most used procedures for finding the location of this minimum are the L-curve [13] (see also [11]) and the generalized cross validation (GCV) [5, 7]. These two techniques are quite general, since they can be applied either to square or to rectangular linear systems. However, both can have deficiencies.

In this paper we shall propose new methods for locating the minimum, under the assumption that the matrix of the linear system is square, nonsingular, and that the right hand side is affected by an error of unknown norm. These methods are based on estimates of the norm of the solution error which are obtained by an extrapolation procedure. Let us mention that extrapolation has already been used for improving the quality of the results obtained by Tikhonov's regularization [3].

Our estimations of the norm of the error are presented in Section 2, where it is shown how they are derived and where their properties are given. Tikhonov's regularization is discussed in Section 3. In Section 4, we explain how our error estimates lead to the choice of a good approximation of the optimal regularization parameter in Tikhonov's method. The results of many numerical experiments are reported in Section 5.

2 Error estimates for linear systems

We consider the $p \times p$ nonsingular system of linear equations

$$Ax = b.$$

Let x^* be any vector. The problem we are addressing is to check whether x^* is a good approximation of the exact solution x. It is well known that the error $e = x - x^*$ is related to the residual $r = b - Ax^*$ by $Ae = r$. Thus, it is not feasible to compute the error from the residual.

If $\|A\|$ or $\|A^{-1}\|$ are known (we note that hereafter the symbol $\|\cdot\|$ denotes the two-norm), the quantities $\|r\|/\|A\|$ and $\|A^{-1}\| \cdot \|r\|$ can be considered as estimates of $\|e\|$, and we have the bounds

$$\frac{\|r\|}{\|A\|} \leq \|e\| \leq \|A^{-1}\| \cdot \|r\|.$$

However, these estimates require the knowledge of the Euclidean norm of A or of its inverse and, moreover, in some cases the bounds can be quite inaccurate.

Although an estimate of any consistent norm of the error was given by Auchmuty [1], we consider it only for the Euclidean norm. In particular, he proved that the quantity

$$\widetilde{e}_3 = \frac{\|r\|^2}{\|A^T r\|}$$

(the notation for this estimate will be made clear below), is an approximation of the 2-norm error $\|e\|$. In fact, the Auchmuty estimate $\widetilde{e}_3$ is a lower bound for $\|e\|$ since, by the Cauchy–Schwarz inequality,

$$(r, r) = \left(A^T r, x - x^*\right) \leq \|A^T r\| \cdot \|e\|.$$

Moreover

$$\frac{\|e\|}{\widetilde{e}_3} \leq \max_{\|y\|=1} \|A^T y\| \cdot \|A^{-1} y\|.$$

This estimate was analyzed in depth by Galantai [6].

More estimates of $\|e\|^2$ were given in [2]. They are denoted by e_i^2, for $i = 1, \ldots, 5$,

$$\begin{aligned} e_1^2 &= c_1^4 / c_2^3 \\ e_2^2 &= c_0 c_1^2 / c_2^2 \\ e_3^2 &= c_0^2 / c_2 \\ e_4^2 &= c_0^3 / c_1^2 \\ e_5^2 &= c_0^4 c_2 / c_1^4 \end{aligned}$$

where

$$\begin{aligned} c_0 &= (r, r) \\ c_1 &= (r, Ar) \\ c_2 &= (Ar, Ar). \end{aligned}$$

It was proved that

$$e_1^2 \leq e_2^2 \leq e_3^2 \leq e_4^2 \leq e_5^2, \tag{1}$$

and lower and upper bounds for $\|e\|^2 / e_i^2$ were given.

However, it was not noticed earlier that these estimates could be gathered into the compact formula

$$e_i^2 = c_0^{i-1} \left(c_1^2\right)^{3-i} c_2^{i-4}, \quad i = 1, \ldots, 5. \tag{2}$$

Moreover, the bounds given in [2] can be rewritten as

$$\frac{\rho^{(3-i)/2}}{\kappa} \leq \frac{\|e\|}{e_i} \leq \kappa \rho^{(3-i)/2}, \tag{3}$$

where $\kappa = \|A\| \cdot \|A^{-1}\|$ and $\rho = c_0 c_2 / c_1^2$. Notice that, by the Cauchy–Schwarz inequality, $\rho \geq 1$.

In order to generalize these formulae, we have to remind how they were obtained. We consider the singular value decomposition (SVD) of the matrix A

$$A = U\Sigma V^T$$

with $UU^T = VV^T = I$, $\Sigma =$diag$(\sigma_1, \ldots, \sigma_p)$ and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p > 0$. Let $u_1, \ldots, u_p$ and $v_1, \ldots, v_p$ denote respectively the columns of the matrices U and V. Then, for any vector y, we have

$$\begin{aligned} Ay &= \sum_{i=1}^{p} \sigma_i (v_i, y) u_i \\ A^T y &= \sum_{i=1}^{p} \sigma_i (u_i, y) v_i \\ A^{-1} y &= \sum_{i=1}^{p} \sigma_i^{-1} (u_i, y) v_i. \end{aligned} \tag{4}$$

Setting $\alpha_i = (u_i, r)$ and $\beta_i = (v_i, r)$, it follows immediately

$$\begin{aligned} c_0 = (r, r) = (U^T r, U^T r) &= \sum_{i=1}^{p} \alpha_i^2 \\ = (V^T r, V^T r) &= \sum_{i=1}^{p} \beta_i^2 \\ c_1 = (r, Ar) &= \sum_{i=1}^{p} \sigma_i \alpha_i \beta_i \\ c_2 = (Ar, Ar) &= \sum_{i=1}^{p} \sigma_i^2 \beta_i^2 \end{aligned} \tag{5}$$

and

$$c_{-1} = (A^{-1} r, r) = (e, Ae) = \sum_{i=1}^{p} \sigma_i^{-1} \alpha_i \beta_i \tag{6}$$

$$c_{-2} = (A^{-1} r, A^{-1} r) = (e, e) = \sum_{i=1}^{p} \sigma_i^{-2} \alpha_i^2. \tag{7}$$

Approximations of c_{-2}, c_{-1}, c_0, c_1 and c_2 can be obtained by keeping only the first term in the summations appearing in the right hand sides of the formulae above. Thus, we shall look for α, β and σ satisfying the interpolation conditions

$$\begin{cases} \alpha^2 = c_0 \\ \alpha\beta\sigma = c_1 \\ \beta^2\sigma^2 = c_2, \end{cases} \tag{8}$$

and then extrapolate for the values -1 and -2 of the index. Thus, c_{-1} and c_{-2} will be respectively approximated by

$$c_{-1} \simeq \alpha\beta\sigma^{-1} \quad \text{and} \quad c_{-2} \simeq \alpha^2\sigma^{-2}.$$

Moreover, we will identify α^2 and β^2 since both furnish a truncated approximation of c_0.

Computing the unknowns α, β and σ from the interpolation conditions (8), which do not have a unique solution, we obtain the 5 estimates given by (2).

The estimates (2) could be extended to any real number ν

$$\|e\|^2 \simeq e_\nu^2 = c_0^{\nu-1} \left(c_1^2\right)^{3-\nu} c_2^{\nu-4}, \quad \nu \in \mathbb{R}. \tag{9}$$

Indeed, replacing the c_i's by their expressions from (8), yields

$$e_\nu^2 = \left(\alpha^2\right)^{\nu-1} \left(\alpha^2\beta^2\sigma^2\right)^{3-\nu} \left(\beta^2\sigma^2\right)^{\nu-4}.$$

After simplification, with the further approximation $\alpha^2 \simeq \beta^2$, we get

$$e_\nu^2 \simeq \alpha^2\sigma^{-2}.$$

Using the inequalities $c_0^2/\kappa^2 \le c_2\|e\|^2 \le c_0^2\kappa^2$ and $0 \le c_1^2 \le c_0c_2$, we obtain bounds generalizing those given in (3)

$$\frac{\rho^{(3-\nu)/2}}{\kappa} \le \frac{\|e\|}{e_\nu} \le \kappa\rho^{(3-\nu)/2}. \tag{10}$$

More estimates can be obtained as follows. We have

$$A^T r = \sum_{i=1}^{p} \sigma_i \alpha_i v_i,$$

and we set

$$\widetilde{c}_2 = (A^T r, A^T r) = \sum_{i=1}^{p} \sigma_i^2 \alpha_i^2.$$

So, the last interpolation condition in (8) could be replaced by

$$\alpha^2\sigma^2 = \widetilde{c}_2,$$

and we obtain the estimates, also valid for any real number ν,

$$\widetilde{e}_\nu^2 = c_0^{\nu-1}(c_1^2)^{3-\nu}\widetilde{c}_2^{\nu-4}, \quad \nu \in \mathbb{R}. \tag{11}$$

Indeed, as above,

$$\widetilde{e}_\nu^2 = (\alpha^2)^{\nu-1}(\alpha^2\beta^2\sigma^2)^{3-\nu}(\alpha^2\sigma^2)^{\nu-4},$$

and the results follows if we assume $\alpha^2 \simeq \beta^2$. Notice that Auchmuty's estimate is recovered for $\nu = 3$.

Bounds similar to those given in (10) could be obtained by replacing ρ by $\widetilde{\rho} = c_0\widetilde{c}_2/c_1^2$. It must be noticed that, when A is orthogonal, $\|e\| = \rho^{(3-\nu)/2}e_\nu = \widetilde{\rho}^{(3-\nu)/2}\widetilde{e}_\nu$. If $\nu = 3$ or if r is collinear to an eigenvector of A, then $e_3 = \widetilde{e}_3 = \|e\|$. For these reasons, these two estimates seem to be the best ones.

Let us now prove an inequality generalizing (1). Formula (9) can be written as

$$e_\nu^2 = \left(\frac{c_0c_2}{c_1^2}\right)^\nu \cdot \left(\frac{c_1^6}{c_0c_2^4}\right) = \rho^\nu e_0^2. \tag{12}$$

Obviously, (11) could be put into a similar form

$$\widetilde{e}_\nu^2 = \left(\frac{c_0\widetilde{c}_2}{c_1^2}\right)^\nu \cdot \left(\frac{c_1^6}{c_0\widetilde{c}_2^4}\right) = \widetilde{\rho}^\nu \widetilde{e}_0^2. \tag{13}$$

But $\rho \geq 1$ and $\widetilde{\rho} \geq 1$, which shows that the estimates e_ν and $\widetilde{e}_\nu$ are increasing functions of ν in $(-\infty, +\infty)$. Thus

$$e_\nu \leq e_{\nu'}, \quad \widetilde{e}_\nu \leq \widetilde{e}_{\nu'}, \quad \nu \leq \nu'.$$

When $\nu \leq 3$, the inequality $c_1^2 \leq c_0\widetilde{c}_2$ can be plugged into (11), thus leading to $\widetilde{e}_\nu^2 \leq c_0^2/\widetilde{c}_2$, and we get a result similar to Auchmuty's [1]

$$\widetilde{e}_\nu^2 \leq \|e\|^2, \quad \forall \nu \leq 3.$$

This inequality has not been proved for $\nu > 3$, nor for e_ν^2.

Thus, by this inequality and (13), and since $\widetilde{e}_\nu^2$ tends to infinity with ν, it exists $\widetilde{\nu} \geq 3$ such that $\widetilde{e}_{\widetilde{\nu}} = \|e\|$. This $\widetilde{\nu}$ is given by

$$\widetilde{\nu} = 2\ln(\|e\|/\widetilde{e}_0)/\ln\widetilde{\rho}. \tag{14}$$

Since e_ν increases from 0 to $+\infty$ when ν covers $(-\infty, +\infty)$, it also exists $\widetilde{\nu}$ such that $e_{\widetilde{\nu}} = \|e\|$. Notice that $c_1^2 = (r, Ar)^2 \leq \|r\|^2 \cdot \|Ar\|^2$, and it follows $e_3^2 \leq \|r\|^2 \leq \|A\|^2 \cdot \|e\|^2$. Thus, if $\|A\| \leq 1$, this $\widetilde{\nu}$ is also greater or equal to 3. Obviously, these values of $\widetilde{\nu}$ cannot be computed in practice.

When A is symmetric and positive definite, $c_{-1} = (e, Ae)$ is the A-norm of the error and, as explained above, it could be approximated by $\alpha\beta\sigma^{-1}$ after solving the system (8). We obtain the following estimates (they have been squared for an easier comparison with the estimates of the norm of the error)

$$(r, A^{-1}r)^2 = (e, Ae)^2 \simeq \widehat{e}_\nu^2 = c_0^{\nu+1}\left(c_1^2\right)^{2-\nu}c_2^{\nu-3}, \quad \nu \in \mathbb{R}. \tag{15}$$

We also have

$$\widehat{e}_\nu^2 = \left(\frac{c_0c_2}{c_1^2}\right)^\nu \cdot \left(\frac{c_0c_1^4}{c_2^3}\right) = \rho^\nu \widehat{e}_0^2 = c_0\rho e_\nu^2 \tag{16}$$

which shows that $\widehat{e}_\nu$ is an increasing function of ν, and that there exists a value of ν such that $\widehat{e}_\nu = (e, Ae)$.

Remark 1 For any $\nu, \mu \in \mathbb{R}$, it also holds $e_\nu^2 = \rho^{\nu-\mu} e_\mu^2$. Ratios of the previous estimates are also estimates of $\|e\|^2$. For example, fixing $\tau, \gamma, \mu \in \mathbb{R}$, we have $e_\gamma^{2\tau}/e_\mu^{2\tau-2} = e_\nu^2$ with $\nu = \tau(\gamma-\mu)+\mu$ for $\tau \geq 1$ (for $\tau = 1$, $\mu = 0$). Similar remarks and relations are also valid for $\widetilde{e}_\nu^2$ and $\widehat{e}_\nu^2$.

3 Tikhonov's regularization

If the system $Ax = b$ is ill-conditioned, Tikhonov's regularization consists of computing the vector x_λ which minimizes the quadratic functional

$$J(\lambda, x) = \| Ax - b \|^2 + \lambda^2 \| Hx \|^2 \tag{17}$$

over all vectors x, where λ is a parameter, and H a given $q \times p$ $(q \leq p)$ matrix. This vector x_λ is the solution of the system

$$(C + \lambda^2 E) x_\lambda = A^T b, \tag{18}$$

where $C = A^T A$ and $E = H^T H$. The vector x_λ is also the least squares solution of the system

$$\begin{bmatrix} A \\ \lambda H \end{bmatrix} x_\lambda = \begin{bmatrix} b \\ 0 \end{bmatrix}. \tag{19}$$

Indeed, multiplying both sides by $[A^T, \lambda H^T]$, leads to (18). Thus, setting $r_\lambda = b - Ax_\lambda$, it holds

$$A^T r_\lambda = \lambda^2 E x_\lambda. \tag{20}$$

Incidentally, let us note that Tikhonov's method can be extended for including several regularization terms [4] as follows

$$J(\lambda, x) = \| Ax - b \|^2 + \sum_{i=1}^{k} \lambda_i^2 \| H_i x \|^2, \tag{21}$$

where λ denotes the multi-index $(\lambda_1, \ldots, \lambda_k)$. The vector x_λ minimizing $J(\lambda, x)$ is the solution of the system

$$\left(C + \sum_{i=1}^{k} \lambda_i^2 E_i \right) x_\lambda = A^T b,$$

where $E_i = H_i^T H_i$, and it can also be written as the least squares solution of a system generalizing (19). In such a case, we have the following generalization of (20) to the multi-parameter case

$$A^T r_\lambda = \sum_{i=1}^{k} \lambda_i^2 E_i x_\lambda. \tag{22}$$

4 Parameter estimation in regularization

In many practical situations, due to the ill-conditioning, if λ is close to zero x_λ is badly computed while, if λ is far away from zero, x_λ is well computed but the norm of the error $x - x_\lambda$ is quite large. For decreasing values of λ, the norm of the error $\|x - x_\lambda\|$ first decreases, and then increases when λ approaches 0. Thus the error, which is the sum of the theoretical error and the error due to the computer's arithmetic, passes through a minimum corresponding to the optimal choice of the regularization parameter. In particular problems, the solution error may present more than one local minimum, or even an interval on which its value is approximately constant, making the search for a good value of the parameter more challenging.

Several methods have been proposed to obtain an effective choice of λ. The most well known are the L-curve [13] and the generalized cross validation (GCV) [5, 7]. The first one consists of plotting in log–log scale the values of $\|Hx_\lambda\|$ versus $\|r_\lambda\|$. The resulting curve is typically L-shaped and the selected value of λ is the one corresponding to the corner of the L. However, there are many cases where the L-curve exhibits more than one corner, or no corner at all. The second method searches for the minimum of a function of λ which is a statistical estimate of the norm of the residual. Occasionally, the value of the parameter furnished by this method may be inaccurate because the function is rather flat near the minimum.

Theoretically, if all the σ_i and all the α_i appearing in formula (7) were known, the value of λ could be estimated by minimizing the right hand side of this formula, that is

$$\sum_{i=1}^{p} \sigma_i^{-2}(u_i, r_\lambda)^2. \tag{23}$$

Such computations were performed, for several values of the regularization parameter, for an orthogonal random matrix with two noise levels, $\varepsilon = 0$ and $\varepsilon = 10^{-8}$ (the two graphs on the left of Fig. 1), and for the matrix *Prolate* [17] (the two graphs on the right of Fig. 1), both of dimension 20. The condition number of this *Prolate* matrix is $5.6 \cdot 10^{+13}$. In both cases the regularization matrix is the identity matrix. The norm of the error vector $x - x_\lambda$ is plotted with respect to the variation of λ as a plain line, while the results obtained by formula (23) are reported in dashed lines.

However, as it could be seen in Fig. 1, even if the σ_i and the α_i can be computed, formula (7) does not produce good results. Moreover, the SVD can only be implemented for quite small matrices and, even in this case, the σ_i and the α_i can be polluted by large rounding errors if the matrix is ill-conditioned. Quite similar problems arise when using (7) for computing the error achieved by the truncated SVD, and also with formula (6) for the energy norm when A is symmetric and positive definite. For these reasons, we will not use (7) for approximating the optimal value of the regularization parameter in Tikhonov's method, but our estimates of the norm of the error.

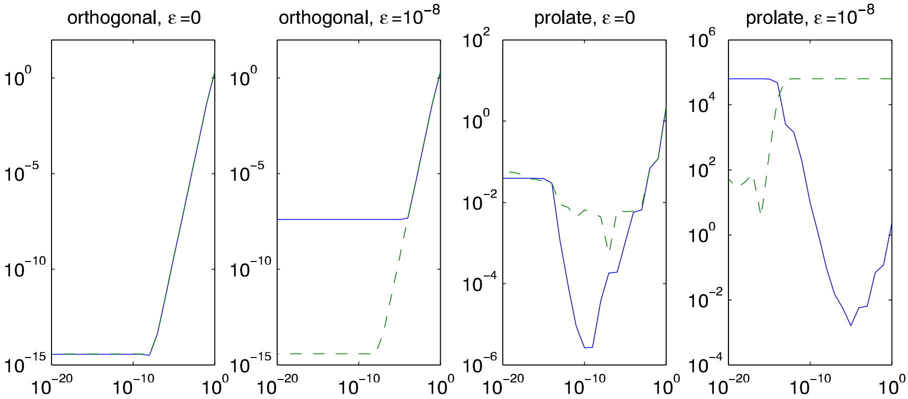


Fig. 1 Exact error (*plain line*) and error computed by formula (23)

There are two questions that have to be answered. Is x_λ well computed? Is x_λ a good approximation of x?

For the first question, we propose the following test. From (20), it holds

$$\frac{\lambda^2 \|Ex_\lambda\|}{\|A^T r_\lambda\|} = 1. \tag{24}$$

So, a regularized solution x_λ could be regarded as accurately computed if this ratio is close to 1, while, on the contrary, either the numerator or the denominator of (24) (or both) could have been strongly affected by the propagation of rounding errors.

For answering the second question, the estimates (9) and (11) could be used. However, due to the ill-conditioning of the problem, the computation of $\widetilde{c}_2 = \|A^T r_\lambda\|^2$, which appears in formulae (11), is not accurate when λ approaches zero. So, according to (20), we will replace $A^T r_\lambda$ by $\lambda^2 E x_\lambda$ in $\|A^T r_\lambda\|$ and in $(r_\lambda, Ar_\lambda) = (A^T r_\lambda, r_\lambda)$, and we finally obtain the error estimates

$$\widetilde{e}_\nu^2 = \|r_\lambda\|^{2\nu-2}(r_\lambda, Ex_\lambda)^{6-2\nu}\|Ex_\lambda\|^{2\nu-8}\lambda^{-4}. \tag{25}$$

Performing the same replacements in (15) or (16) for the A-norm of the error, we get

$$\widehat{e}_\nu^2 = \|r_\lambda\|^{2\nu+2}(r_\lambda, Ex_\lambda)^{4-2\nu}\|Ex_\lambda\|^{2\nu-6}\lambda^{-4}. \tag{26}$$

Let us remark that $(r_\lambda, Ex_\lambda) = (Hr_\lambda, Hx_\lambda)$, which avoids computing the matrix E and, in several cases, leads to a more stable procedure.

Contrarily to the more general estimates (9) and (15), which are always valid (that is for any direct or iterative numerical method for solving a system of linear equations), formulae (25) and (26) are specially adapted to Tikhonov's regularization. So, they should lead to better numerical results. Testing the equality in (24) is also only valid for Tikhonov's regularization.

For theoretical purpose, let us now show that the estimate $\widetilde{e}_3$ is sharp, that is there exist regularization matrices for which it is exact (even if such a matrix cannot be computed in practice). To this end, we take $H = v_k v_k^T$, for $1 \leq k \leq p$. We remind that, since $Av_k = \sigma_k u_k$ and $A^T u_k = \sigma_k v_k$, we have

$$x = \left(I + \lambda^2 C^{-1} v_k v_k^T\right) x_\lambda$$

$$C^{-1} v_k = \frac{1}{\sigma_k^2} v_k,$$

$$r_\lambda = \frac{\lambda^2}{\sigma_k}(v_k, x_\lambda) u_k,$$

$$A^T r_\lambda = \lambda^2 (v_k, x_\lambda) v_k$$

$$x - x_\lambda = \frac{\lambda^2}{\sigma_k^2}(v_k, x_\lambda) v_k.$$

So, it follows

$$\|x - x_\lambda\| = \frac{\lambda^2}{\sigma_k^2}|(v_k, x_\lambda)|.$$

On the other hand

$$\|r_\lambda\|^2 = \frac{\lambda^4}{\sigma_k^2}(v_k, x_\lambda)^2$$

$$\|A^T r_\lambda\| = \lambda^2 |(v_k, x_\lambda)|,$$

and we finally obtain, as claimed,

$$\|x - x_\lambda\| = \frac{\|r_\lambda\|^2}{\|A^T r_\lambda\|} = \widetilde{e}_3.$$

For the same choice of H, it could be checked that the estimate $\widehat{e}_3$ is exact, that is $\widehat{e}_3 = (x - x_\lambda, A(x - x_\lambda))$. We stress again that this choice of the regularization matrix, although not a practical one, is only an example to show that there are cases for which the estimates are exact.

The preceding considerations are also valid in the multi-parameter case (21), which can be treated similarly.

Remark 2 Another way of approximating the error could be based on preconditioning. We have $x - x_\lambda = \lambda^2 C^{-1} E x_\lambda$. Let $P \simeq A$ be a preconditioner for A, and replace in the preceding expression C by $Q = P^T P$ which is a preconditioner for $A^T A$. Then, an approximation of the error could be obtained by one of the following formulae

$$x - x_\lambda \simeq \lambda^2 Q^{-1} E x_\lambda = Q^{-1} A^T r_\lambda.$$

Obviously, Q can be directly taken as a preconditioner of $A^T A$ without building P first. However, the numerical experiments we performed show that these estimates are very sensitive to the choice of the preconditioner.

5 Numerical experiments

We performed many numerical experiments to ascertain the performance of our estimates in the selection of the regularization parameter both in Tikhonov's method and in a regularizing iterative method (namely, the conjugate gradient algorithm). These experiments were executed with Matlab 7.4 [14] on an AMD64 computer running Debian Linux. The computation of the regularized solutions and the selection of the parameter by the L-curve and GCV methods were performed by the Regularization Tools [10]. The Structured Matrices Toolbox [16] was used for the solution of large scale Toeplitz systems by iterative methods.

We developed some Matlab functions, which automatize the application of our estimates to Tikhonov regularization and iterative methods. The software, which is available upon request, includes a few simple functions which implement formulae (9), (11), (15), (25) and (26). These functions may be used in Tikhonov regularization either when working on a grid, or if the parameter is discrete, i.e. in iterative methods. Moreover, we developed a function specialized for Tikhonov's method, which searches for the minimum of formula (25) by an adaptive algorithm, computing the regularized solution either by employing the SVD in (17) or solving the normal system (18) by the conjugate gradient method, depending on the user's choice. The details of this algorithm, which aims to minimize the number of evaluations of $\widetilde{e}_\nu$, each of which implying the solution of a linear system, will be described in a forthcoming paper.

To construct each test linear system, we selected various matrices from the Regularization Tools and the `gallery` function of Matlab (see list in Table 6). For each matrix we computed the right hand side corresponding to a solution chosen from a set of vectors with different degree of regularity, described in Table 1. If the noise level is ε, this means that each component of the right

Table 1 Solution vectors

given	default solution for problems from [10], `shaw` solution for the others
ones	$x_i = 1$
linear	$x_i = \frac{i}{p}$
parabola	$x_i = \left((i - \lfloor \frac{p}{2} \rfloor) / \lceil \frac{p}{2} \rceil\right)^2$
sin2pi	$x_i = \sin \frac{2\pi(i-1)}{p}$
linear+sin2pi/4	$x_i = \frac{i}{p} + \frac{1}{4} \sin \frac{2\pi(i-1)}{p}$
step	$x_i = 0, i \leq \lfloor \frac{p}{2} \rfloor$ $x_i = 1, i > \lfloor \frac{p}{2} \rfloor$

Table 2 Results for Example 1

ν	0	1	2	3	4	5
e_ν	$6.33 \cdot 10^{-19}$	$8.15 \cdot 10^{-18}$	$1.05 \cdot 10^{-16}$	$1.35 \cdot 10^{-15}$	$1.74 \cdot 10^{-14}$	$2.24 \cdot 10^{-13}$
$\widetilde{e}_\nu$	$6.28 \cdot 10^{-19}$	$8.10 \cdot 10^{-18}$	$1.04 \cdot 10^{-16}$	$1.35 \cdot 10^{-15}$	$1.74 \cdot 10^{-14}$	$2.25 \cdot 10^{-13}$
$\widehat{e}_\nu$	$2.51 \cdot 10^{-32}$	$3.23 \cdot 10^{-31}$	$4.16 \cdot 10^{-30}$	$5.36 \cdot 10^{-29}$	$6.90 \cdot 10^{-28}$	$8.89 \cdot 10^{-27}$

hand side is perturbed by adding a normal random variable (null mean value and unitary variance) scaled by the quantity ε.

To measure the quality of the results we used the 2-norm error $\|e\| = \|x - x^*\|$ of an approximate solution x^* with respect to the exact solution x.

5.1 Example 1

To show how our estimates approximate the error, we first considered a well-conditioned system of dimension 15 with a random matrix, whose condition number is 1378.3, and we assumed $x = (1, \ldots, 1)^T$ as the solution vector. Solving it by Gaussian elimination we obtained $\|e\| = 5.548 \cdot 10^{-14}$, $|(e, Ae)| = 9.486 \cdot 10^{-29}$, and $\widetilde{\nu} = 4.45$ [see (14)]. For $\nu = 0, 1, \ldots, 5$, we get the results shown in Table 2. Notice that, although the matrix is not symmetric, $\widehat{e}_\nu$ is a very good estimate of $|(e, Ae)|$.

5.2 Example 2

In this section we give two examples to illustrate the behavior of our estimates in Tikhonov regularization, in comparison to the L-curve and the GCV. The first example is the Wing test problem from [10] with $p = 40$. The condition number is about $1.6 \cdot 10^{19}$. The Tikhonov functional (17) is minimized by means of the SVD, and the regularization matrix H is the identity matrix.

The results are shown in Fig. 2, where the thick line gives the Euclidean norm of the error with respect to the variation of λ, the dashed one is $\widetilde{e}_3$, and the dash-dotted line corresponds to the graph of the GCV function multiplied

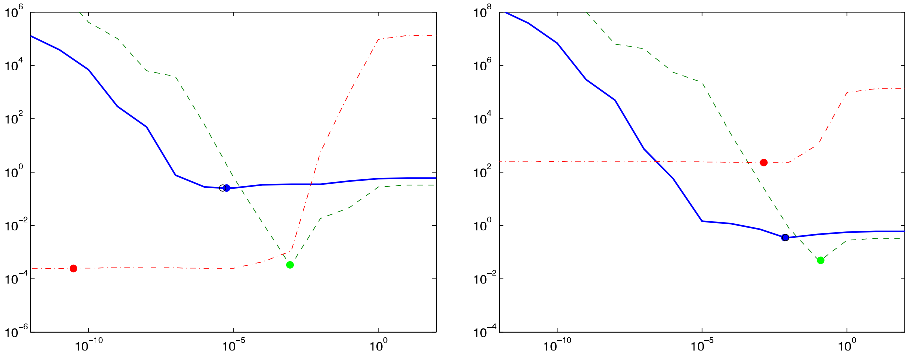


Fig. 2 Wing example: left $\varepsilon = 10^{-6}$, right $\varepsilon = 10^{-3}$

Table 3 Parameters and errors for the Wing example

		Optimal	$\widetilde{e}_3$	L-curve	GCV
$\varepsilon = 10^{-6}$	λ	$5.8 \cdot 10^{-6}$	$9.0 \cdot 10^{-4}$	$4.3 \cdot 10^{-6}$	$3.0 \cdot 10^{-11}$
	$\|e\|$	$2.5 \cdot 10^{-1}$	$3.4 \cdot 10^{-1}$	$2.5 \cdot 10^{-1}$	$2.8 \cdot 10^{4}$
$\varepsilon = 10^{-3}$	λ	$7.2 \cdot 10^{-3}$	$1.2 \cdot 10^{-1}$	$7.7 \cdot 10^{-3}$	$1.4 \cdot 10^{-3}$
	$\|e\|$	$3.5 \cdot 10^{-1}$	$4.7 \cdot 10^{-1}$	$3.5 \cdot 10^{-1}$	$5.6 \cdot 10^{-1}$

by a factor 10^{10}. On each curve, the minimum is indicated by a bullet. Where it is visible, a circle indicates the value of the error coming out from the L-curve. The graph on the left shows the results for a noise level $\varepsilon = 10^{-6}$, while, on the right, it corresponds to $\varepsilon = 10^{-3}$. Table 3 displays the values of the parameters and the errors furnished by the various methods. In the table, the "optimal" parameter is the one which minimizes the error.

In this case, the L-curve gives very good results, while the GCV function is extremely flat, resulting in a false minimum when $\varepsilon = 10^{-6}$, which produces a very large error. The estimate $\widetilde{e}_3$ tends to overestimate the optimal parameter, but this is generally less dangerous than underestimating it, and the errors are acceptable.

In the second example we used the Pascal matrix with the solution coming out from the Shaw problem from [10], $p = 20$, $\varepsilon = 10^{-6}, 10^{-3}$ and $H = I$. The condition number is roughly $1.2 \cdot 10^{20}$. The results are in Fig. 3 and Table 4. In this case the GCV gives good results with both noise levels, while the L-curve returns an off-scale parameter with a large error. Our technique still overestimates the optimal parameter, but produces a good approximation of the solution.

5.3 Example 3

Here we consider a particular example on which our method totally fails. Since the construction of our estimate is based on truncating the SVD of the cofficient matrix to just one term and then extrapolating, we built a matrix with many large singular values.

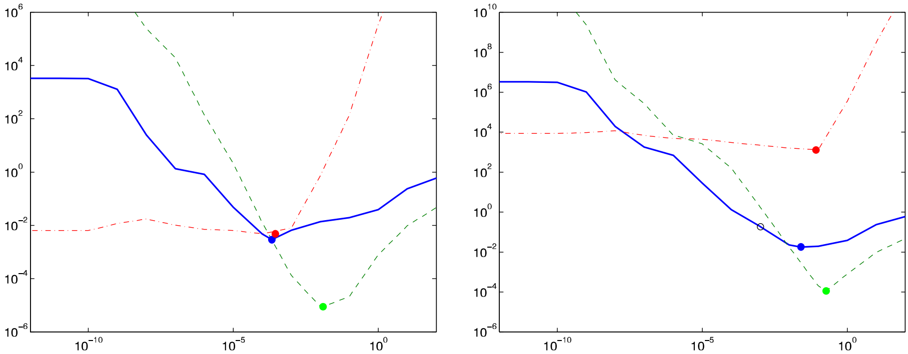


Fig. 3 Pascal example: *left* $\varepsilon = 10^{-6}$, *right* $\varepsilon = 10^{-3}$

Table 4 Parameters and errors for the Pascal example

		Optimal	$\widetilde{e}_3$	L-curve	GCV
$\varepsilon = 10^{-6}$	λ	$2.1 \cdot 10^{-4}$	$1.2 \cdot 10^{-2}$	$2.7 \cdot 10^{9}$	$2.8 \cdot 10^{-4}$
	$\|e\|$	$2.9 \cdot 10^{-3}$	$1.4 \cdot 10^{-2}$	4.4	$3.1 \cdot 10^{-3}$
$\varepsilon = 10^{-3}$	λ	$2.5 \cdot 10^{-2}$	$1.9 \cdot 10^{-1}$	$1.0 \cdot 10^{-3}$	$8.4 \cdot 10^{-2}$
	$\|e\|$	$1.8 \cdot 10^{-2}$	$2.0 \cdot 10^{-2}$	$1.8 \cdot 10^{-1}$	$1.9 \cdot 10^{-2}$

So, we let $p = 20$ and

$$\sigma_i = \begin{cases} 10, & i = 1, \dots, 7, \\ e^{-.08(i-8)^2}, & i = 8, \dots, p. \end{cases}$$

Then, fixing a random vector w with unitary norm, we construct a Householder matrix $U = I - 2ww^T$, and we set

$$A = U \mathrm{diag}(\sigma_i) U^T.$$

The right hand side of the system corresponds to the solution $x_i = .05$, $i = 1, \dots, p$. We took $\varepsilon = 10^{-3}$ and $H = I$. The results are given in Fig. 4 and Table 5.

As it can be seen, the L-curve and the GCV both give good results, while $\widetilde{e}_3$ has the global minimum for a very small value of the parameter and a local minimum near the optimal value.

5.4 Example 4

To better investigate the performance of our estimates in Tikhonov regularization, compared with the L-curve and the GCV, we consider 12 matrices (1st column of Table 6), 7 different solutions (Fig. 1), 2 dimensions (20 and 100), 3 regularization matrices (identity, discretization of first and second

Fig. 4 Results for Example 3

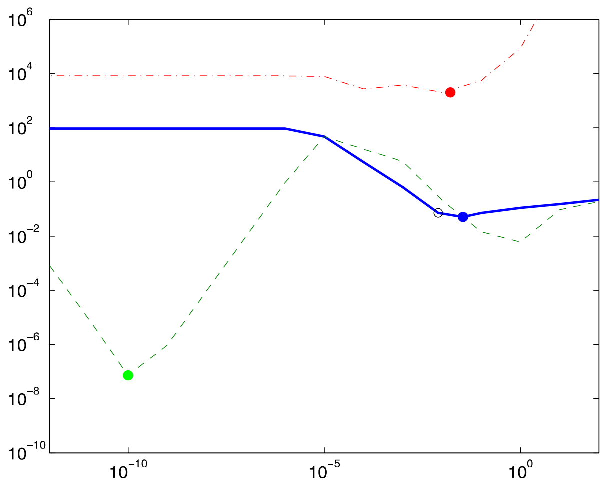


Table 5 Parameters and errors for Example 3

	Optimal	$\widetilde{e}_3$	L-curve	GCV
λ	$3.4 \cdot 10^{-2}$	$1.1 \cdot 10^{-10}$	$8.0 \cdot 10^{-3}$	$1.6 \cdot 10^{-2}$
$\|e\|$	$5.1 \cdot 10^{-2}$	$9.4 \cdot 10^{1}$	$7.2 \cdot 10^{-2}$	$6.0 \cdot 10^{-2}$

derivatives), 4 noise levels ($\varepsilon = 10^{-6}, 10^{-4}, 10^{-2}, 10^{-1}$), and 5 realizations of the random noise. So, we have a total number of 10,080 experiments, 840 for each test matrix.

For each parameter estimation method, we measure the quality of the result by computing the ratio of the norm of the error corresponding to the estimated parameter divided by the norm of the minimal error. So, this ratio is always greater or equal to 1, and it equals 1 only when the estimated parameter furnishes the optimal error.

The parameter coming out from our estimate is computed by the above mentioned adaptive minimization algorithm in the interval $[10^{-30}, 10^{2}]$. The values of λ for the L-curve and the GCV are obtained using the functions `l_curve` and `gcv` from [10]. The optimal parameter is estimated by minimizing the function

$$\psi(\rho) = \|x - x_\lambda\|, \quad \text{with } \lambda = 10^{\rho},$$

by the `fminsearch` function of Matlab, using as a starting point the logarithm of the parameter furnished either by our estimate, or by the L-curve, or by the GCV, which produces the lowest error. The regularized solution is then computed by the `tikhonov` function from [10], which minimizes (17) by applying the SVD.

In Table 6, we list what we consider the *failures* and *severe failures*, i.e. we give the number of ratios larger than 10^2, and those greater than 10^4, obtained by $\widetilde{e}_3$, the L-curve, and the GCV, respectively. The condition numbers of the matrices of dimension 20, as computed by the `cond` function of Matlab,

Table 6 *Failures* of methods for estimating λ, vs. test matrices

Matrix	Cond.	$\widetilde{e}_3$		L-curve		GCV	
		$> 10^2$	$> 10^4$	$> 10^2$	$> 10^4$	$> 10^2$	$> 10^4$
Baart	$2.0 \cdot 10^{17}$	0	0	149	121	123	68
Heat(1)	$1.0 \cdot 10^{20}$	26	0	283	190	39	0
Hilbert	$3.1 \cdot 10^{18}$	0	0	156	131	52	40
Ilaplace(3)	$9.2 \cdot 10^{30}$	2	0	231	218	45	24
Lotkin	$2.1 \cdot 10^{19}$	0	0	168	138	42	33
Moler	$1.7 \cdot 10^{13}$	223	48	235	49	173	42
Pascal	$1.2 \cdot 10^{20}$	509	280	548	383	298	280
Phillips	$4.0 \cdot 10^{03}$	247	20	108	36	31	0
Prolate	$5.6 \cdot 10^{13}$	0	0	149	105	114	84
Random	$2.8 \cdot 10^{02}$	80	8	202	44	141	12
Shaw	$9.9 \cdot 10^{15}$	0	0	113	96	63	8
Wing	$1.7 \cdot 10^{19}$	0	0	245	201	78	56
Total		1,087 11%	356 4%	2,587 26%	1,712 17%	1,199 12%	647 6%

Table 7 *Failures* of methods for estimating λ, vs. regularization matrices

H	$\widetilde{e}_3$		L-curve		GCV	
	$> 10^2$	$> 10^4$	$> 10^2$	$> 10^4$	$> 10^2$	$> 10^4$
I	311	160	390	130	372	144
D_1	386	39	939	643	402	248
D_2	390	157	1,258	939	425	255
Total	1,087	356	2,587	1,712	1,199	647

are displayed in the second column. A realistic estimation of the condition numbers of the matrices of dimension 100 is not possible, and so we don't give them.

It is evident that, on this set of examples, the method we propose is more robust than the L-curve and has a performance slightly better than the GCV, especially for what regards the *severe failures*. We note that all the matrices used in this test are severely ill-conditioned, with the exception of the random and the Phillips matrices. The reason for including these test problems is to investigate how the methods behave in the presence of a well conditioned (or mildly ill-conditioned) matrix.

In Tables 7 and 8, the same results are displayed disaggregated with respect to regularization matrices and to noise levels. It appears that the estimate $\widetilde{e}_3$ is not particularly sensitive to these aspects. On the contrary, as it is obvious given its statistical foundation, the GCV performance improves as the noise level increases.

We observe that, in these tests, using the expression of $\widetilde{e}_3$ given by formula (25) is fundamental. In fact, if we repeat the tests using formula (11) to estimate the error, the number of failures grows from 1,087 to 7,166.

5.5 Example 5

For large matrices, it is not possible to use the SVD to minimize the Tikhonov functional (17). As a consequence, the classical implementation of the GCV method is not applicable in these cases, since it also involves the use the SVD. However, these difficulties may be overcome by an algorithm, based on

Table 8 *Failures* of methods for estimating λ, vs. noise level

ε	$\widetilde{e}_3$		L-curve		GCV	
	$> 10^2$	$> 10^4$	$> 10^2$	$> 10^4$	$> 10^2$	$> 10^4$
10^{-6}	343	117	724	399	500	226
10^{-4}	302	78	560	302	320	174
10^{-2}	197	78	520	378	202	129
10^{-1}	245	83	783	633	177	118
Total	1,087	356	2,587	1,712	1,199	647

Gaussian quadratures, which devises lower and upper bounds for the GCV function and is particularly suited for large scale problems [8].

In this example, for $\lambda = 10^{-10}, 10^{-9}, \ldots, 10^{-1}, 1$, we solve the system (18) by the conjugate gradient algorithm (CG) up to convergence. The matrix is *Prolate* of dimension 10,000 with the *parabola* solution (see Fig. 1) and a noise level of 10^{-4}.

In Fig. 5, we show the results obtained letting $\nu = 2$ in formula (11) (dashed line) and in formula (25) (plain line). The thick line represents the exact error. It is immediate to observe that both formulae identify correctly the minimum of the error.

5.6 Example 6

Since the conjugate gradient itself has a regularizing effect on the solution of a system $Ax = b$, here we use our estimates for stopping the iterations of CG.

We take the *Gaussian* matrix [15] of dimension 10,000, with a *ones* solution, and a noise level of 10^{-4}. Its asymptotic condition number, as defined in [15], when the parameter ρ equals 0.01, is $1.0 \cdot 10^{214}$. To solve this system we use the `pcg` function of Matlab and the `smt` toolbox [16] which implements optimized storage and fast arithmetics for Toeplitz and circulant matrices.

On the left of Fig. 6, we show the error in thick line, and the estimates $\widetilde{e}_\nu$ for $\nu = 1, \ldots, 5$, versus the iterations. The minima are indicated. They almost all arise at the same iteration. On the right, we display the error (thick plain line), the A-norm error $(e, Ae)^{1/2}$ (thick dashed line), and the estimates $\widetilde{e}_3$ (thin plain line) and $(\widehat{e}_3)^{1/2}$ (thin dashed line). The best error is $3.8 \cdot 10^{-2}$, and it is attained at iteration 46. Our estimates find a minimal error of $4.0 \cdot 10^{-2}$ at iteration 52. For the L-curve, a new algorithm for finding its corner was recently developed in [12]. In the special case of a discrete regularization parameter, it gives better results than the subroutine used in [10], and, in our case, it found the corner at iteration 93 with an error of $6.3 \cdot 10^{-1}$.

Fig. 5 Tikhonov/CG: Prolate matrix

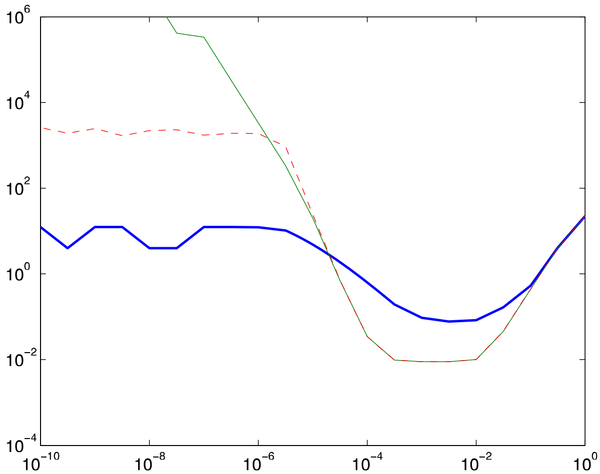
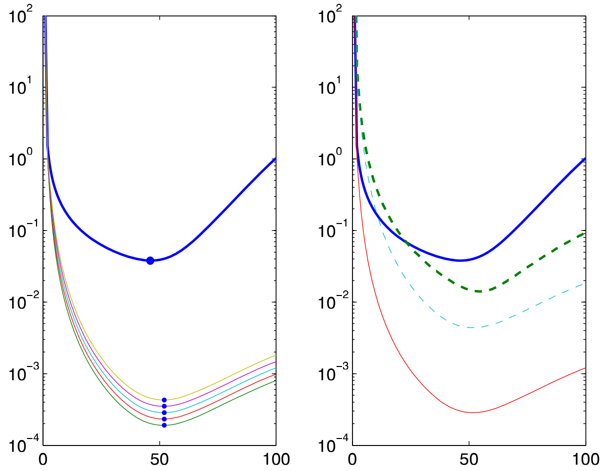


Fig. 6 Regularizing CG: Gaussian matrix



On this example, we see that the A-norm of the error is better estimated than its L^2-norm, maybe because $\widetilde{e}_3$ does not take into account the symmetry and the positive definiteness of the matrix A, while $\widehat{e}_3$ does.

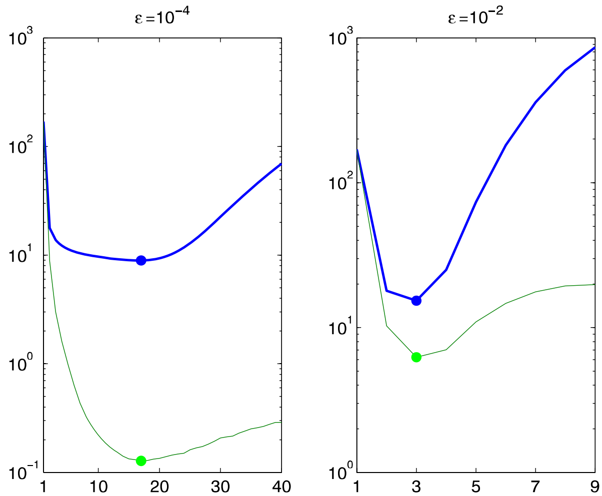
Let us mention that estimations of the norm of the error specially adapted to CG were obtained by Golub and Meurant [9].

5.7 Example 7

Finally, let us solve an image deblurring problem by CG. The size of the image is 256×256, and so the dimension of the system is $256^2 = 65{,}536$.

We initially apply a Gaussian blur to a test image, displayed on the left of Fig. 8, and contaminate it with a noise at level 10^{-4} and 10^{-2}. Figure 7 reports

Fig. 7 CG: deblurring problem



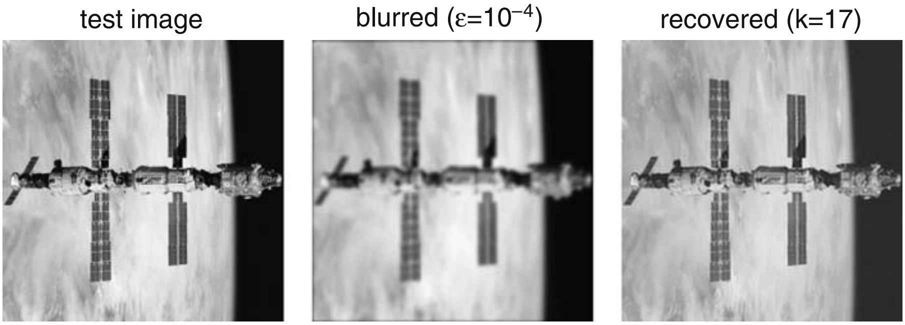


Fig. 8 Images for the deblurring problem

the graph of the error (thick lines) and of e_3 (thin lines) for the two noise levels. Minima are indicated by a bullet.

The observed image and the reconstruction, for $\varepsilon = 10^{-4}$, are shown in Fig. 8.

6 Conclusions

In the first part of this paper, we extend the estimates given in [1] and [2] for the norm of the error of any approximate solution of a system of linear equations. We also give some properties of these estimates.

In the second part of the paper, we apply these estimates to the search for the best regularization parameter λ in Tikhonov's regularization method. Computing this best value is a difficult problem. To this end, when the norm of the error is unknown, two methods are essentially used: the L-curve, and GCV. Each of them has its own drawbacks. Sometimes, the L-curve does not exhibit a clear corner (corresponding to the optimal choice of the parameter). On the other hand, the GCV function does not always present a clear minimum.

If an iterative method presenting a regularization effect is used for solving the system, then the error initially becomes to decrease and then, due to the propagation of initial errors, it increases. So, an efficient stopping criterion has to be used.

If the SVD of the matrix of the system is available, in theory, the norm of the error could be computed by formula (7). However, if the matrix is ill-conditioned, the SVD is subject to rounding errors, and, if it is large, it may not even be possible to obtain it. Motivated by these considerations, in this paper we propose other techniques, based on various estimates of the norm of the error. Of course these techniques, as any other, do not lead to an all-purpose method. Nevertheless, our numerical experiments highlight that it is a quite trustworthy procedure. Moreover, it is easy and cheap to implement. So, it could either be an effective alternative to other techniques, or supplement them.

Finally, we note that our method, unlike GCV and L-curve, is not applicable to rectangular linear system. Furthermore, its performance in the solution of large scale problems should be compared with the algorithm proposed in [8] for the computation of the GCV. Both these themes will be addressed in a forthcoming paper.

Acknowledgements C. Brezinski would like to acknowledge the financial support of the University of Cagliari for an invited professorship stay during which this work was done, and to thank S. Seatzu and G. Rodriguez for their warm hospitality. The authors thank the two referees for their helpful remarks which led to a substantial improvement in the content and the presentation of the paper.

References

1. Auchmuty, G.: A posteriori error estimates for linear equations. Numer. Math. **61**, 1–6 (1992)
2. Brezinski, C.: Error estimates for the solution of linear systems. SIAM J. Sci. Comput. **21**, 764–781 (1999)
3. Brezinski, C., Redivo-Zaglia, M., Rodriguez, G., Seatzu, S.: Extrapolation techniques for ill-conditioned linear systems. Numer. Math. **81**, 1–29 (1998)
4. Brezinski, C., Redivo-Zaglia, M., Rodriguez, G., Seatzu, S.: Multi-parameter regularization techniques for ill-conditioned linear systems. Numer. Math. **94**, 203–228 (2003)
5. Craven, P., Wahba, G.: Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross validation. Numer. Math. **31**, 377–403 (1979)
6. Galantai, A.: A study of Auchmuty's error estimate. Comput. Math. Appl. **42**, 1093–1102 (2001)
7. Golub, G.H., Heath, M., Wahba, G.: Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics **21**(2), 215–223 (1979)
8. Golub, G.H., von Matt, U.: Generalized cross-validation for large-scale problems. J. Comput. Graph. Stat **6**(1), 1–34 (1997)
9. Golub, G.H., Meurant, G.: Matrices, Moments and Quadrature with Applications. Princeton University Press, Princeton (to appear)
10. Hansen, P.C.: Regularization tools: a Matlab package for analysis and solution of discrete ill-posed problems. Numer. Algorithms **6**, 1–35 (1994)
11. Hansen, P.C.: Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion. SIAM, Philadelphia (1998)
12. Hansen, P.C., Jensen, T.K., Rodriguez, G.: An adaptive pruning algorithm for the discrete L-curve criterion. J. Comput. Appl. Math. **198**, 483–492 (2006)
13. Hansen, P.C., O'Leary, D.P.: The use of the L-curve in the regularization of discrete ill-posed problems. SIAM J. Sci. Statist. Comput. **14**, 1487–1503 (1993)
14. Matlab ver. 7.4, The MathWorks, Inc., Natick, MA (2007)
15. van der Mee, C.V.M., Seatzu, S.: A method for generating infinite positive self-adjoint test matrices and Riesz bases. SIAM J. Matrix Anal. Appl. **26**, 1132–1149 (2005)
16. Redivo-Zaglia, M., Rodriguez, G.: `smt`: a Matlab structured matrices toolbox (2007) (in preparation)
17. Varah, J.M.: The prolate matrix. Linear Algebra Appl. **187**, 269–278 (1993)